International Journal of Engineering & Technology, 7 (4) (2018) 3478-3484

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET
doi: 10.14419/ijet.v7i4.15882
Research paper

Survey on big data security in various applications

Ambica Sethy [1] *, Abhishek Ray [1]

[1] *School of Computer Engineering, Kalinga Institute of Industrial Technology Deemed to be University, Bhiabneswar, India*
**Corresponding author E-mail: ambica.sethy@gmail.com*

Abstract

The present era has become a digital world, which has been serving the mankind in each and every sector. Due to this, a lot of varied data is used and stored across the digital world. To handle this tremendous varied data, the concept of big data plays a vital role. Data and big data are two sides of a single coin. Digitization has been made possible along with the growth of data through Big data. This leads to creation of many security concerns in Big Data applications. This paper presents a review on some of the existing security and privacy issues on different big data application domains and triggered some new additional V's like visualization plus vulnerability, which has been correlated with the security and privacy issues on those domains. This paper also focuses on that application area which has not taken security and privacy issues of big data into consideration. We have also found some new issues that could become possible threats for the protective data in various applications.

Keywords: *Applications; Big Data; Privacy; Security; Vulnerability.*

1. Introduction

Big data is basically associated with the globalization of the present era, where digitization has taken its prominent role to play in it. Big Data is a new approach for enhancing decision making by processing large volume, velocity, variety of information. When the size of data exceeds beyond the ability of typical database warehouse, then the concept of Big Data came into the picture to capture, store, manage and analyze such data. Real world databases are so much bulky in size that processing becomes unwieldy for handling such databases in terms of size, speed and processing time.

Basing on some real life data utilization such as in banking sector, where numerous data are stored regarding the customers, staffs and transactions, it is difficult to analyze these voluminous data which is also seen in the social networking sites like Facebook, Twitter, What's app. Many more domains has been storing their data with a belief to access its availability with it's need which has been challenging role to be followed [1] [2]. Big Data problem include capture, storage, dissemination, search, analytics, and visualization. Some of the key decisive power provided by Big Data can be categorized as defining needs, revenue generation, and decision strategy, better services and identifying new trends. With the increase of data size, its vulnerability increases according to that various tools plus techniques can be developed who make it secure; where the intruders have proved themselves more advanced in security breaching. The magnanimous infrastructure of cloud and huge inter-cloud data migration, a variety of data sources and firms, and immense data attainment invite security vulnerabilities [4].

Big Data along with being beneficial towards the surplus amount of data, it has challenges and issues as well. One of the challenging concerns to be preserved with Big Data is security and privacy issues; some of them we are concerned such as confidentiality, integrity, availability, monitoring and auditing, key management, and that can be considered in Big Data application domains such as Biotech, defense, ERP/CRM, Finance/Banking, Geo-science weather, mobile telecom, marketing, business analytics. Taking Big Data security into concern, we have noticed some big security breaching phenomena/techniques which are used in different Big Data application domains to break its security. Further, we have detected some additional privacy issues from the existing application domains and tried to frame some solution for it[5].

2. Preliminaries

This section discusses all the basic concepts of security and privacy challenges in Big Data plus with its elements.

2.1. Big data

Big Data is a buzz word which is now whispering everywhere in every sector. Big data is expressed in terms of data relating to bigger volume, heterogeneous sources with distributed and decentralized control, that can handle complex relationship between data. It constitutes both structured and unstructured data that grows large so faster that they are not manageable by traditional relational database systems or conventional statistical tools. It has been better demonstrated through a theorem called as HACE Theorem (Heterogeneous, Autonomous, Complex and Evolving) [7]. This theorem has been assumed upon a giant elephant which is treated as big data in this context. The giant elephant needed to be analysed by 6 blind men to give their perception on it, as shown in Fig.1 which is correlated with "Big Data" [6].

Copyright © 2018 Ambica Sethy, Abhishek Ray. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

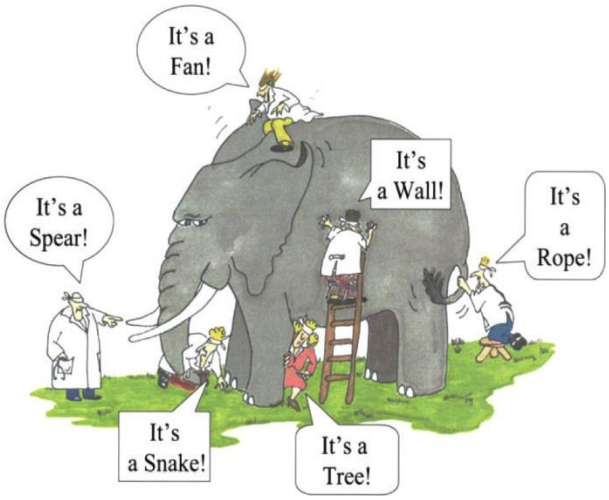


Fig. 1: Example of V's.

The main goal is, how big data comes into the picture in front of the society and observe based on reaction of the society. The blind persons had the different point of view about elephants such as a tree, a big fan, a huge wall, a solid spear, a snake or a rope due to lack of actual knowledge of reality. If it is taken as partial information about a particular data such as all the blind persons analyze only the parts of total output which is collected during their analysis. This can be used to explain the important key elements of big data by assuming the elephant as growing size link with volume, different pose as velocity and its every body part as variety. Partial information may create biases of the actual information. This example can have a clear outline of Big data for laymen. Exploring ``Big Data " in this situation can be considered as combining information from various diversified sources (blind men). Big Data Analytics is a process of: Collecting, Organizing Analyzing of large sets of data to Discover patterns and other valuable information.

3. Security and privacy issues provoked by 8vs of big data

According to (Gartner IT Glossary) "Big data is high volume, high velocity, and high variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making ". Here V's of big data are summarized to co-relate security and privacy with V's as shown in Fig.2.

3.1. Volume

It defines the key characteristic of big data. Volume is the amount of data generated by organizations or individuals. Data are generated by the organization in the sense per day how much data are generated by the company which are established for our social development which is connected with web world we can take that organization like social media such as facebook, twitter, LinkedIn or we can say news channels, other TV channels. The sources may be typical internal or internal-external data sources, Some typical Internal data sources are File systems, SQL, NoSQL, archives of Scanned documents, paper archives, records etc and internal-external data source can be Sensor Data, Social media, Business apps, Media, DOCs. One of the key concerns for privacy is storage and use of data [15]. Volume indicates how to store the huge amount of data; it should need a big platform to store data. The big volume of data increases the risk of information leakage. So this element is directly associated with security and privacy [8].

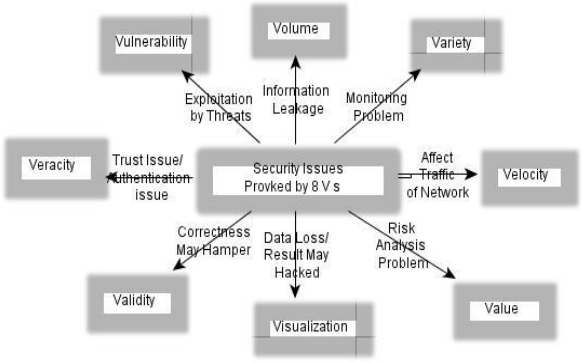


Fig. 2: 8 V's of Big Data.

3.2. Variety

It deals with Different types of data and data types, which is originated from different sources. This data is generally categorized into three types: Structured data, Semi-structured data, and unstructured data.

i) Structure data: The data which has predefined format it can be easily stored in a traditional database which can be represented in a row and column format [8].

ii) Semi-structured data: It is just a part of structured data but it cannot fit into each and every data model. It is difficult for us to store this type of data in a traditional database. This type of data can store in a file which consists of a number of tags, for example, XML, JSON, etc.

iii) Unstructured data: This type of data does not have any predefined formatted structure. Each and every entity are not mapped with every attribute ex- image, video etc.

Variety is defined by its number of sources with natures of data. Here the data comes from different sources and the data are of a different format, so data format problem occurred and to manage, store them become impossible. The unformatted data resulted in an invalid solution for any infrastructure use and it may affect the privacy of data. Problems like data monitoring plus data auditing arises due to massive data, for this reason security issues like confidentiality, authentication or integrity can get introduced by default.

3.3. Velocity

It represents the data rate at which data is generated and spread. It shows the uninterrupted and high frequencies of data. The security and privacy problem comes in this 'V' also and the most common issues are information security when the data are generated at the high speed. Fast growing and iterating data severely affect the traffic of the network, which can slow down the whole process[8]. This slowing down can cause a server crash, which can result in failure of validation mechanism by which hackers can gain access to the resources which they are not allowed to access. Besides, the hacker can launch advanced persistent threats (APT) more easily, while it is hard to be detected by the traditional protection strategy.

3.4. Value

Value is defined such a way suppose the industry has some important data than for the risk analysis or we can say finding some new idea from the existing information we need to analyse them and after the analysis, what type of result we will get that term indicates the value of big data. This value is summarizing much more thing for the betterment and it derives business value from the data also. The output is produced from a large number of data sets and this amount of data may cause the security problem, if any information is leaked by the attacker this will hamper the future of an organization, because of this v also create security as well as the privacy issue.

3.5. Visualization

It is a way of representing the processed data. Once the resulting process is over, then the task is to describe or represent the result. Why we are using the visualization tools? The content of result does not explore the basic idea plus essential thing about the actual raw data. Here also the data and results are not safe, once the resulting process is complete so many accidents may happen, if the result is hacked by any attacker everything will lose. With the help of data visualization software expose all information which input is given that input may be a bar, a line or may be any text data. So here for a large amount of data if we are not conscious it may occur privacy issues. To encode visual objects contained in graphics like points, lines or bar for communicating with data or informations the techniques used is Data Visualization. The primary objective is to communicate information undoubtedly and proficiently with users, for which it has been a important part of data science. Visualization provides instant outline of the processed data for better decisive power, and is basically used for exploring new data set.

3.6. Validity

Ensuring the purpose before sharing information, life of the importance of any data. Whether the obtained data is correct or not. The parameter you are considering. Correctness is defined by availability one of the security characteristics.

3.7. Veracity

It refers to the accuracy and trustworthiness of data or we can say the uncertainty of data. This is related to authentication security problem [20]. It is strongly connected to veracity, one of the five Vs.

3.8. Vulnerability

Identified or unidentified flaws in any system protocols. This is a serious weakness of the security program or we can say this is gap made by the experts which they can't full fill till now, this is also related with big data and comes under a serious challenge for every expert in data science. Always it is exploited by threats to gain authorized access to a particular resource. From 8 Vs this V is also facing so many challenges [4]. Hackers have hacked driver-less Google cars by jamming the steering wheel and showcased the risk of demanding bounty from passengers if the science of hacking is made public.

4. Security characteristics

"You can have security and not have Privacy, But you cannot have Privacy without Security" Tim Mother.

In near recent stages, computer system damages, generally due to viruses such as malware, ransom-ware (WannaCry, Petya). Organizations by taking the benefit of this, are used to circulate these type of contents to advertise their antivirus software for selling them. As this is the era of big data, this type of malware activities have been reduced but the security and privacy fear are still hunting us. With companies moving forward to implementing big data concepts the fear of privacy leak increases even more. Traditional security, trust plus privacy methods has been affecting due to the new inventions of IT tycoon. Because of different forms of attacks in the application side and attacks with terrible effect in the hardware component, security has become the most difficult task in Big data [10].

i) Confidentiality: It is the main feature while talking about security, which is mainly used during dealing with sensitive data's. When the data is found to be confidential during storing and processing, it gets permission in gathering process. Information should secure between sender and receiver, ensuring secrecy of information.
ii) Authentication: Ensuring only legitimate users can access the services without any false identity; this is a key characteristic of security. This service provided the purity of the transaction.
iii) Availability: Service must be available to the authorized user (instead of non-repudiation in security service).
iv) Integrity: Here one question will occur regularly i.e. "How do I ensure that the data which are stored that is really stored my data without tampering with it?". Data integrity is the legality and superiority of data. This is also one of the key characteristics of the security.
v) Non-repudiation: Taking digital security into concern, one of the characteristics called non-repudiation provides proof of the integrity and origin of data. Authentication guarantees genuineness with high certainty by the help of non-repudiation. Repudiation means denying. Non-repudiation is to ensure that someone cannot deny something [9]. It has become one of the key security characteristics for data transfer, where either sender or receiver's authenticity or identity cannot be refused by either of them.
vi) Access Control: Access control is a security skill that can use to Prevention of unauthorized access to a resource and services [5]. Most important thing is that services must be accessible to the authorized user, having any technical knowledge or not. Technically access control can divide two types, physical and logical. Physical access control identified the boundary and IT assets, logical access control represents computer networks, system files, and data.[search security]

5. Big data application areas

Different application areas of different sectors have been benefited by using big data for their analysis purpose. Big data had helped them in analysing their data to uncover some important patterns and more valuable information [21]. In this way, big data is helpful in every application areas, which are externally or internally connected with our society. In Healthcare sector, big data is playing a vital role for their patients as well as doctors, by keeping the detailed analysis of the individual profile. Big data has also become popular in public sectors, private sector is technically connected with big data. Kinsey Global Institute has come up with the importance of Big data in some main areas [3]:-

5.1. Biotech

It has helped in improving medical decision support system by analysing disease patterns and getting them benefited in finding the present solution as well as easy and quick access to the data and methods in future, without overloading work pressures.

5.2. Public sector

Public sectors like banking, telecommunication, government sector, military, police, infrastructure (public roads, bridges, tunnels, water supply, sewers, electrical grids, telecommunications)by directly or indirectly helping in decision making, giving idea about customer motive and interest, risk analysis with solutions, innovating new products with services, development of automated systems to suppress risk.

5.3. Marketing

This sector has a personal relationship with big data from its origin, Big Data allows businesses to personalize each customer's experience and it even lets them calculate consumer behaviour which is the key point of marketing. It helps in analysing customer's mood over a purchase of products, their interest in web-based

markets. Big data improves product activity design, helps in-store behavior analysis, and optimizes labour inputs, logistics optimization, new business ideas are innovated.

5.4. Industrialization

This sector and big data are interconnected, for enhanced demand forecasting, supply chain planning, sustaining sales, industrial production operations, web exploration based applications.

5.5. Private sector

Private sectors helps to provide easy information as well as brief knowledge about smart routing, geo-targeted advertising, metropolitan development, innovative trade models.

6. Literature survey

The concept of Big Data security and privacy has been defined in many ways, in this section, we quantitatively and qualitatively analyse the 45 papers according to their areas. There are many applications domains of big data which are comes under these fields such as industrialization, Public sector, private sector and many more, they are not given that much importance to system reliability and security. However, very few prior studies focus on security of the ``information and communication technology''. It is critically important as it is the base of all the applications.

Demchenko et al. [23] introduces the Scientific Data Life cycle Management (SDLM) model which consists of all important stages by reflecting the basics of data management in modern e-Science, along with proposed the SDI generic architecture model, which gives the details for building inter operable information and explain how these models can be naturally implemented using modern cloud based communication. It also proposed the important infrastructure factors for Big Data, which refers to different technical communities for the requirements on data management, security and access control. But it could include non-repudiation technique for giving more security to the scientific data infrastructure; it could help to keep the communication more authenticate with identity proof.

Chandra et al. [12] described the current health care security issues with its challenges in big data environment. One of the most important factors to concern is, due to increase adoption of cloud technologies, threats of breaches and leakage of confidential data also increases. So, by using Big Data environment, it is trying to secure the healthcare system, basically the patient health record. Here, different Big Data security approaches such as Health Care Big Data and Internet of Things, Health Care Big Data and Cloud, Securing Patient Health Records in Big Data and Cloud Environments plus health Care and Big Data Analytics has been used. There is a requirement to secure the sensitive health care data to maintain the privacy and integrity of the delicate data from the middleman and Malicious software.

Patil et al. [11] represented the big data in different way by presenting the state-of-the-art issues of security and privacy in the healthcare industry. To decrease the threats of security and privacy issues, more focus on the scientifically burst through computational, storage and communication capabilities could be strived.

Rao et al. [13] presented some usable security solutions to prevent the exploitations of the potential big data in healthcare area. The solutions are carried out in a highly synchronized environment. According to the author, the solutions should be developed in such way that it must ensure the protection of analysis and its Frameworks. In order to enhance the quality of health care, one of a Act in the U.S. called The Health Information Technology for Economic and Clinical Health (HITECH), has enforced for the adoption of electronic health records (EHRs) [14]. New Data Sources (NDS) project refers to patient-generated health data, consumer device data, fitness data, and data from social media, but it has not taken big data along with privacy and security into primary concern [15].

Wang et al. [18] discuss some ethical implications of big data privacy and privacy preserving technologies related to record linkage, synthetic data generation, and genomic data privacy in biomedicine. It need to deal with many problems and rising challenges, because we believe good solutions to diminishes privacy risks in biomedical research require a combined effort from different areas. Ancy et al. [38] proposed a method to determine the condition of a patient as normal or kidney failure patient and handle easily large volumes of clinical datasets namely, RBFNN (Radial Basis Function Neural Network) with classifier algorithm with the use of parameters. By the help of this method it tells all the stages of kidney failure patient and treatment. Most significant issue is that security and privacy, we need to technically handled these data sets for the future use with maintain the secrecy of datasets use.

Bates et al. [16] has presented different ways for addressing privacy part of the health care systems and also developed some support systems for research on analytics, by specifying some use cases in the field of high-cost patients, readmissions, triage, decompensation, adverse events, and treatment optimization for diseases affecting multiple organ systems using Big Data analysis to minimize cost. The organization always need to perform some basic analysis and changes need to be implemented to improve the system along with cost minimization. Feng et al. [17] proposed a digital device array called Raspberry pi (RP) for data collection by taking big data security issues into account in a health care situation. The device can play a very vital role in the field of Digital Forensic Science, since a systematic approach is used in the device to establish security issues in the health and smart city areas. RP device has that much potential to carry out in any poor configuration if security threats arises in future. Health care system is unable to deal with todays cons pirated environment because of the lack of using big data analytics, its security and privacy [19].

Camargo et al. [24] proposed a big data analytics system to measure citizens' awareness through Twitter social network. This could be made possible by taking government permission with the help of some set of Big Data tools. To distinguish the security characteristics in tweets, a machine learning algorithm is needed to be learned. The final reports are described in Heat Map to check whether the tweets are secure or not. And most significant thing is that government should provide Big Data tools to utilize all the technology of Big Data. As per Lu and Wu [25] the most vital key patterns for creating a healthy and balanced new media ecological environment are the safeguard methods to provide and consume the sensitive data. And also how efficiently these data are used to provide safety information. The difficulty of the new media big data security research is that understanding the new media content and identify the new media ideas.

Shen et al. [26] proposed a fog-assisted mobile crowd sensing architecture and two privacy preserving crowd sensing schemes for crowd sensing applications. The method is based on an additive homomorphic encryption algorithm and a bitwise-XOR homomorphic encryption algorithm. By the help of these algorithms the service subscriber collects the statistical and accurate data, without any background details and link between individual data from each applicant. But for security we have to secure these architecture model and its data content of each participant.

Kshetri [27] enlightened the costs, profits, uses, and future scopes associated with organizations and studied how big data are related to privacy, security of consumer welfare. Along with all the benefits, effects of big data on the way of security and privacy of complexity, vulnerability and technological knowledge's is also discussed, which vary from consumers to consumers. Besides all these, how Big Data raise any ethical issues, misuses has also been pointed out. There is a need to develop our environment, to stop the exploitation of consumer, ignore their interests. In case of privacy issues, consumers are often unknown regarding their lackness of up-to date information. Hence, every organization needs to

have a firm-level big data policy for our better and knowledgeable future.

Zhang et al. [28] proposed an on-line and status-aware approach called SafeDrive where labelled data are not required like others. SafeDrive statistically develops a state graph (SG) as a activities model in off line. To depart an anomaly from the SG, SafeDrive classifies some segments. SafeDrive is accomplished to identify various types of driving anomalies. By the help of this approach we can alert rivers to correct their driving behaviours. Instead of these we have to deeply analyse the driver driving behaviour with complete road data such as videos, road network knowledge, and vehicles information for quick and clear identification of anomalies with all description.

Jain et al. [29] summarize various types of weather forecasting applications belonging to different domains using Big Data and defined technical challenges facing while taking advantages from these domains. One problem is, perception of any weather reporter plays important role, since a small wrong perception will cause severe loss in every sector. To remove these types of threats, we have to identify an optimal solution for these types of disputes, which can save lives. The wrong perceptions may occur due to any privacy and security issues of systems.

Lin et al. [22] proposed a framework that is based on a longitudinal operational cooperation network model, of terrorist organizations and the sources are derived from an open source database. A quantitative analysis is done for the vibrant change of cooperation relations, for which the framework is divided into two sub levels like network-level and node-level. In network-level, it has found the dynamic change process of network using topological structure and in node-level; how a position has changed using centrality analysis has been detected. And consider a valuable technique based on Big Data from open source intelligence. Global Terrorism Database (GTD) is used in this analysis for the dataset where the terrorist incidents are recorded with brief information around the world from 1970 to 2014. All the updated information regarding 50 dangerous terrorist organizations can be accessible in an on-line platform called Big, Allied, and Dangerous (BAAD). ORA is also used to study the structural dynamic change in the network. But here the cooperation networks are loosely connected which may cause the serious global terrorist attacks. We have to focus on security and privacy of the networks to secure that network.

Yan et al.[30] proposed a reference value to intelligent safety supervision in the big data platform and mentioned four important technologies such as integration, analysis, processing and display technologies to fulfil the need of intelligent safety supervision. To develop the construction of intelligent safety supervision platform and find the value of smart safety supervision, big data analysis technology should combine all technical advantages and safety supervision system applications required. Janssen et al. [31] illustrate the government's role, availability for citizens control, impact on society, opportunities to the public, challenges to transparency and privacy in the Big and Open Linked Data (BOLD). Since BOLD creates a great impact in the society and public organization, there is a huge need of the fundamental and applied research in the area. Though privacy and transparency are fundamental principle, they should take the complex nature in the role and establish a relationship with other variables as they are not at all well understandable.

Puthal et al. [32] proposed a new authenticated key exchange scheme that is Dynamic Prime Number Based Security Verification (DPBSV) scheme, which is based on symmetric key cryptography and random prime number generation that renewed dynamically. Main aim is to provide well organized and fast security verification for big data streams. The common shared key updates both source sensing device and data stream manager (DSM). To assure end-to-end security and maintain data qualities DSM have to always verify. Some applications need to remove redundant data and get original data for stream data analysis. To improve the efficiency of symmetric key encryption towards more efficient security aware big data streams, need to investigate for the further new approach. Hardy et al. [33] judge the benefits and risks of declaring government data as open data and classified the challenges that faced by the Australian government due to released the public personal data as a open data into the public domain. Open data is a data which is accessible for free and which can be accessed by anybody and re-used for any purpose. The Government should focus on developing more secure and trustworthy techniques to permit the release of government data, that may hamper any public personal life.

Zhao et al. [34] explained the concepts, characteristics and data processing steps of electric power Big Data on the basis of opportunities and new security challenges faced when big data is use for analysis. It needs to make stronger the traditional information security to make the big data as a new driving force of the era.

Tonni et al. [35] proposed a microaggregation technique to provide data anonymity in case of its type. Microaggregation technique is used to protect a dataset through anonymity, which may expose a person's identity. To secure these data sets used two types of algorithms in the microaggregation methods namely an evolutionary attribute grouping algorithm and Huffman data compression algorithm. Therefore, these algorithms can be applied to secure big data in CPS applications that require continues data access, and prevent data disclosure. Improving the algorithm performance we need to added and sorted a new cumulative attribute frequency into the priority queue.

Terzi et al. [36] studied network anomaly and attack detection with a new supervised anomaly detection approach on Apache Spark cluster in Azure HDInsight in the big data analysis. The results are visualized as 3D by dimension reduction with principal. Component Analysis to detect policy violations, outlier's malicious traffic flow easily. In network traffic, most of the flows are normal that's why it may affect the detection of anomalies, so for the better results we need to more data, anomalies, efficient algorithm and platforms. Palaiokrassas et al. [37] presents a recommendation service based on Neo4j graph data base and a modelling approach, which is applied in the perspective of smart cities application and leverages the potential of big data. Challenges and issues of Neo4j also studied and validate. The service is applied by taking the open big data and user generated data through Smartphone applications. But for the upcoming advance generation of graph data bases, we need to use larger and more complex datasets, further leveraging on the effectiveness of the social networking services and experiment with new techniques.

Zaki et al. [39] presents a concise summary on big data and e-mail security plus analyse two case studies to observe the security threats for big data. Case study such as Enron email dataset to investigate the security challenges of big data in email and the next is take 35 under graduate students to observe how the phishing email generate through a users intention and break the security system. Finally results showed that big data analysis helps phishers to understand the behaviour of email users. Therefore we concluded here through big data all organizations, every individual persons suffers by using of phishing. It can destroy everything of a company, because it may attacks all the communities of a company. There need to propose a framework and a set of guidelines for the prevention of such big data security threat in e-mail communication. Cockcroft et al.[40] presents a snapshot of big data academic research in information systems to identify themes in accounting, finance research and practice. From 2007-2016, 47 journal papers are used to analysis of accounting, finance and information systems. After the survey identify a taxonomy of themes which is presented as a conceptual matrix. Themes are used as concepts and the matrix identifies where they appear. Key point is that very few research works has done on privacy and security of big data in accounting and finance. Increased research in these areas which will lead to improvements in industry practices, and opportunities for cross-disciplinary research.

Martin et al. [41] used a conceptual framework based on gossip theory which are directly and indirectly links with customer vulnerability. Prevent the negative performance effects on marketers and customers relationship this framework is used. Three case studies are performed to show transparency and control in firms'

data management performance, which can make smother the negative effects of customer data vulnerability. Mainly problem is that, here consider not only about the firm's but also the customers data whatever the express, their feeling, how much trust they have upon the company and after this type of event study of data security breaches affects badly to public companies also confirms the negative effects. We should try to recover from these negative events by the help of vulnerability encouraging events to restore philanthropic features of customer relationships.

7. Performance evaluation

As per the literature survey Table 1 shows the mapping with usage of big data in different application domains in terms of percentage and come to found new issues that could be encountered those areas.

Table 1: Evaluation Mapping

Domains/ Potentiality	Usage of Big Data (%)	Prioritization of Security and Privacy (%)	Issues encountered (after big data usage) (%)
Healthcare [11]	70	60	20
Telecome, Defence & Electrcal[34]	30	20	10
Market Research [41]	80	60	30
Industrialization [27]	40	20	10
Finance and Insurance [40]	20	10	10

In the table1 all domains specified has been categorized into two sectors that is public and private sector. In the public sector Healthcare has given much more priority to big data for analyzing their work along with security plus privacy concern. But other public sector like telecoms, electrical, defense has not applied that much big data into their concern. In private sector one of the domain such as market research which include amazon, flipkart, snapdeal, medlife, big basket using Big Data analytics as their key database handler for enhance market business. The area gives priority to security plus privacy for their consumer satisfaction. But, some domains like finance and insurance has still not that much aware about the importance of big data in the growth of the organization. In various sector more issues are decreasing after the use of big data, but in some cases security issues still arises, due to lack of importance given to the security plus privacy factors.

8. Conclusion and future work

More we move towards digitizing the world with IOT, we need to find the meaning of data and store only relevant information for the future use. This data could be helpful, while installing sensors in each appliance, social media networking, and industrial sector. In order to find out the important pattern from these raw data, Big Data analysis helps readily. This paper has surveyed different application domains working in the areas of big data, prior to the security, privacy and its challenges. We came to the conclusion that, many domains have not used big data in their analysis purpose and had not taken security and privacy issues to their concerns. From the literature survey, we got some security issues in every application domain. Therefore we need to focus on those application areas which has not concentrate on security and privacy issues of big data. One of the issue can be protected data is being used by unauthorized people, which may affect the near future. The protected data after being utilized should be deleted from the database. But some time due to human error, it may not be removed permanently from that system/database, which can be accessed by the hacker [42]. Our future work is based on overcoming these problems in our real world application areas.

References

[1] (2015) [Online]. Available: http://www.gartner.com/it-glossary/bigdata.

[2] Yang Q (2015). Introduction to the ieee transactions on big data. IEEE transactions on big data. (1):2-15.

[3] Sagiroglu S, Sinanc D (2013). Big data: A review. Proceedings of the IEEE International Conference on InCollaboration Technologies and Systems (CTS), (pp. 42-47).

[4] Chin WL, Li W, Chen HH (2017). Energy big data security threats in IoT-based smart grid communications. IEEE Communications Magazine. 55(10):70-5. https://doi.org/10.1109/MCOM.2017.1700154.

[5] Nelson B, Olovsson T (2016). Security and privacy for big data: A systematic literature review. Proceedings of the IEEE International Conference on InBig Data (Big Data), (pp. 3693-3702).

[6] Joshi N, Kadhiwala B (2017). Big data security and privacy issues—A survey. IEEE. InPower and Advanced Computing Technologies (i-PACT), 2017 Innovations in 2017 Apr 21 (pp. 1-5).

[7] Wu X, Zhu X, Wu GQ, Ding W (2014). Data mining with big data. IEEE transactions on knowledge and data engineering. (1):97-107.

[8] Ye H, Cheng X, Yuan M, Xu L, Gao J, Cheng C (2016). A survey of security and privacy in big data. Proceedings of the IEEE 16th International Symposium on InCommunications and Information Technologies (ISCIT), (pp. 268-272).

[9] Sahi MA, Abbas H, Saleem K, Yang X, Derhab A, Orgun MA, Iqbal W, Rashid I, Yaseen A (2018). Privacy Preservation in e-Healthcare Environments: State of the Art and Future Directions. Ieee Access. 2018; 6: 464-78. https://doi.org/10.1109/ACCESS.2017.2767561.

[10] Pearson S, Benameur A (2010). Privacy, security and trust issues arising from cloud computing. Proceedings of the IEEE Second International Conference on InCloud Computing Technology and Science (CloudCom), (pp. 693-702).

[11] Patil HK, Seshadri R. Big data security and privacy issues in healthcare. InBig Data (BigData Congress), 2014 IEEE International Congress on 2014 Jun 27 (pp. 762-765). IEEE.

[12] Chandra S, Ray S, Goswami RT (2017). Big Data Security in Healthcare: Survey on Frameworks and Algorithms. Proceedings of the IEEE 7th International Conference on InAdvance Computing Conference (IACC), (pp. 89-94).

[13] Rao S, Suma SN, Sunitha M (2015). Security solutions for big data analytics in healthcare. Proceedings of the IEEE Second International Conference on InAdvances in Computing and Communication Engineering (ICACCE), (pp. 510-514).

[14] Wang S, Bonomi L, Dai W, Chen F, Cheung C, Bloss CS, Cheng S, Jiang X (2016). Big data privacy in biomedical research. IEEE Transactions on Big Data. (1):1-.

[15] Asghar MR, Lee T, Baig MM, Ullah E, Russello G, Dobbie G (2017). A Review of Privacy and Consent Management in Healthcare: A Focus on Emerging Data Sources. arXiv preprint arXiv:1711.00546.

[16] Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G (2014). Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Affairs. 33(7):1123-31. https://doi.org/10.1377/hlthaff.2014.0041.

[17] Feng X, Onafeso B, Liu E (2015). Investigating big data healthcare security issues with Raspberry Pi. Proceedings of the IEEE International Conference on InComputer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), (pp. 2329-2334).

[18] Wang S, Bonomi L, Dai W, Chen F, Cheung C, Bloss CS, Cheng S, Jiang X (2016). Big data privacy in biomedical research. IEEE Transactions on Big Data. (1):1-.

[19] Gao Y, Wang Z, Ji C, Xiao P, Qin J, Li Z (2017). Design and implementation of a mobile-health call system based on scalable kNN query. Proceedings of the IEEE 19th International Conference on Ine-Health Networking, Applications and Services (Healthcom), (pp. 1-6).

[20] Patel K, Patel B, Mishra M, Patel N (2017). Privacy issues in big data. Proceedings of the IEEE 2nd International Conference on InConvergence in Technology (I2CT), (pp. 259-264).

[21] Khanduja V, Arora A, Garg S (2017). Applications of big data in real world: It's not what you know. It's what you do with what you know. Proceedings of the International Conference on InComputing, Communication and Automation (ICCCA), (pp. 159-163). IEEE.

[22] Lin Z, Sun D, Li Z, Tang M (2017). Research on the dynamic change of terrorist organization cooperation from the big data per-

spective. Proceedings of the IEEE 2nd International Conference on InBig Data Analysis (ICBDA), (pp. 363-367).
[23] Demchenko Y, Grosso P, De Laat C, Membrey P (2013). Addressing big data issues in scientific data infrastructure. Proceedings of the IEEE International Conference on InCollaboration Technologies and Systems (CTS), (pp. 48-55).
[24] Camargo JE, Torres CA, Martínez OH, Gómez FA (2016). A big data analytics system to analyze citizens' perception of security. Proceedings of the IEEE International Conference on InSmart Cities Conference (ISC2), (pp. 1-5).
[25] Lu ZW (2016). Research about New Media Security Technology Base on Big Data Era. Proceedings of the IEEE 14th Intl Conf on InDependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech), (pp. 933-936).
[26] Shen W, Yin B, Cheng Y, Cao X, Li Q (2017). Privacy-preserving mobile crowd sensing for big data applications. Proceedings of the IEEE International Conference on InCommunications (ICC), (pp. 1-6).
[27] Kshetri N (2014). Big data′ s impact on privacy, security and consumer welfare. Telecommunications Policy. 38(11):1134-45. https://doi.org/10.1016/j.telpol.2014.10.002.
[28] Zhang M, Chen C, Wo T, Xie T, Bhuiyan MZ, Lin X (2017). SafeDrive: online driving anomaly detection from large-scale vehicle data. IEEE Transactions on Industrial Informatics. (4):2087-96. https://doi.org/10.1109/TII.2017.2674661.
[29] Jain H, Jain R (2017). Big data in weather forecasting: Applications and challenges. Proceedings of the IEEE International Conference on InBig Data Analytics and Computational Intelligence (ICBDAC), (pp. 138-142).
[30] Chen Y, Li X (2017). Research on big data application in intelligent safety supervision. Proceedings of the IEEE 2nd International Conference on InBig Data Analysis (ICBDA), (pp. 473-477).
[31] Janssen M, van den Hoven J. Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy?
[32] Puthal D, Nepal S, Ranjan R, Chen J (2017). A dynamic prime number based efficient security mechanism for big sensing data streams. Journal of Computer and System Sciences. 83(1):22-42. https://doi.org/10.1016/j.jcss.2016.02.005.
[33] Hardy K, Maurushat A (2017). Opening up government data for Big Data analysis and public benefit. Computer Law & Security Review. 33(1):30-7. https://doi.org/10.1016/j.clsr.2016.11.003.
[34] Zhao J, Wang Y, Xia Y (2016). Analysis of Information Security of Electric Power Big Data and Its Countermeasures. Proceedings of the IEEE 12th International Conference on InComputational Intelligence and Security (CIS), (pp. 243-248).
[35] Tonni SM, Rahman MZ, Parvin S, Gawanmeh A (2017). Securing big data efficiently through microaggregation technique. Proceedings of the IEEE 37th International Conference on InDistributed Computing Systems Workshops (ICDCSW), (pp. 125-130).
[36] Terzi DS, Terzi R, Sagiroglu S (2017). Big data analytics for network anomaly detection from netflow data. Proceedings of the IEEE International Conference on InComputer Science and Engineering (UBMK), (pp. 592-597).
[37] Palaiokrassas G, Charlaftis V, Litke A, Varvarigou T (2017). Recommendation service for big data applications in smart cities. Proceedings of the IEEE International Conference on InHigh Performance Computing & Simulation (HPCS), (pp. 217-223).
[38] Ancy S, Cornelius K (2017). A predictive model for kidney failure E-health. Proceedings of the IEEE Third International Conference on InSensing, Signal Processing and Security (ICSSS), (pp. 497-502).
[39] Zaki T, Uddin MS, Hasan MM, Islam MN (2017). Security threats for big data: A study on Enron e-mail dataset. Proceedings of the IEEE International Conference on InResearch and Innovation in Information Systems (ICRIIS), (pp. 1-6).
[40] Cockcroft S, Russell M (2018). Big data opportunities for accounting and finance practice and research. Australian Accounting Review. https://doi.org/10.1111/auar.12218.
[41] Martin KD, Borah A, Palmatier RW (2017). Data privacy: Effects on customer and firm performance. Journal of Marketing. 81(1):36-58. https://doi.org/10.1509/jm.15.0497.
[42] Reardon J, Basin D, Capkun S (2013). Sok: Secure data deletion. InSecurity and Privacy (SP), 2013 IEEE Symposium on 2013 May 19 (pp. 301-315). IEEE.